Spatial Consistency in 3D Tract-Based Clustering Statistics

Matthan Caan[1,2], Lucas van Vliet[2], Charles Majoie[1], Eline Aukema[1], Kees Grimbergen[1], and Frans Vos[1,2]

[1] Department of Radiology, Academic Medical Center, University of Amsterdam, NL
[2] Quantitative Imaging Group, Delft University of Technology, NL
m.w.a.caan@tudelft.nl.

Abstract. We propose a novel technique for tract-based comparison of DTI-indices between groups, based on a representation that is estimated while matching fiber tracts. The method involves a non-rigid registration based on a joint clustering and matching approach, after which a 3D-atlas of cluster center points is used as a frame of reference for statistics. Patient and control FA-distributions are compared per cluster. Spatial consistency is taken to reflect a significant difference between groups. Accordingly, a non-parametric classification is performed to assess the continuity of pathology over larger tract regions. In a study to infant survivors treated for medulloblastoma with intravenous methotrexate and cranial radiotherapy, significant decreases in FA in major parts of the corpus callosum were found.

1 Introduction

Over the past decade Magnetic Resonance Diffusion Tensor Imaging (MR-DTI) has been used to characterize the local white matter structure in the human brain. Typically, properties derived from the tensor, such as the Fractional Anisotropy (FA) [1], are involved to study differences between diseased subjects and normals.

It may be observed that not all the tensor information (represented by orientation and shape) is used in the latter analysis. For instance, connectivity information is neglected as tensor shape properties are studied on a voxel- or region-of-interest-basis. Alternatively, fiber tracts are used as a subspace in which properties are averaged. However, high variation in white matter anatomy exists along white matter tracts [2]. Not surprisingly, analysis is currently focusing on profiles along tracts. For instance, in a study to Amyotrophic Lateral Sclerosis (ALS), it was recently found that only parts of the corticospinal tract were affected by the disease [3]. Here, a Mann-Whitney U-test was performed to assess local spatial consistency.

Prior to such inter-subject comparison of tract-based features, correspondence between the tracts is to be achieved. Anatomical variability across subjects needs to be accounted for in a non-rigid manner. Preferably, correspondence is not only obtained along, but also in the plane normal to the local tract orientation.

D. Metaxas et al. (Eds.): MICCAI 2008, Part I, LNCS 5241, pp. 535–542, 2008.

In case of small patient and control groups, diffusion properties need to be spatially combined to ensure sufficient sensitivity for discrimination. This may be achieved by a statistical description, e.g. some mean property such as FA over the tract and the corresponding standard deviation in case of a unimodal distribution. Clearly, such averaging goes at the cost of losing spatial information. A more specific approach is to employ the correlation in the data, after which it is found which voxels do or do not contribute to an observable difference between patients and controls [4].

Relevant related work includes affine matching of fibers [5], such that diffusion properties are represented along tracts. Inter-subject comparison is not facilitated by either approach, though. Joint probabilistic clustering and point-by-point mapping achieves arc-length tract representations [6]. Tract-based morphometry also uses such an arc-length representation, based on matched tract regions to a chosen prototype. Subsequently, it is used for detecting white matter differences between groups of subjects [7]. Neither do these methods achieve non-rigid tract correspondence, nor is an unsupervised partitioning of tracts facilitated. Tract based spatial statistics projects non-rigidly registered volumes onto a skeletonization of the FA [8], but tensor orientation information is not used in the alignment, nor in the analysis.

We propose a framework for tract-based analysis that partitions tracts in an unsupervised manner. The method involves a non-rigid registration based on a joint clustering and matching approach, after which a 3D-atlas of cluster center points is used as a frame of reference for statistics. Underlying shape coherence of trajectories is captured in anisotropic cluster shapes. Patient and control FA-distributions are compared per cluster. Spatial consistency is taken to reflect a significant difference between groups. Accordingly, a non-parametric classification is performed to assess the continuity of pathology over larger tract regions.

2 Method

2.1 Feature Selection

The analysis starts with selecting fiber tracts of interest both in patients and controls. We perform a full brain tractography using the FACT algorithm in DTIStudio software [9], after which ROIs are placed that, combined by logical operators, define a sub-selection of fibers, here the corpus callosum.

2.2 Registration

Initially, the fibers need to be brought into spatial correspondence. The concept of point set matching is applied to points along the fiber tracts in order to do so [10].

Suppose that in total, P datasets are to be aligned. Each dataset $p \in \{1, \ldots, P\}$, is represented by sampling the selected fiber tracts, yielding point-set consisting of N^p points $X^p = \{x_i^p, i = 1 \ldots N^p\}$. For each point set, a set of K cluster center points is defined, denoted by $C^p = \{c_a^p | a = 1 \ldots K\}$. An atlas cluster point

set $\mathcal{Z} = \{z_a | a = 1 \dots K\}$ is defined that is to explain the cluster centers over all point sets. By means of a deterministic annealing approach, the registration process gradually refines from global to local matching. The initial temperature T_{init} and final temperature T_{final} are a priori defined. T is lowered after each iteration: $T_{ind+1} = \alpha T_{ind}$, with $0 < \alpha < 1$.

A mixture of Gaussians models the density of the point set, in which the membership variable m_{ai}^p indicates the degree to which a point "feature" x_i^p belongs to cluster center point c_a^p. The membership m_{ai}^p is now defined as

$$m_{ai}^p = \frac{q_{ai}^p}{\sum_{a=1}^K q_{ai}^p}, \forall a, i, \tag{1}$$

with

$$q_{ai}^p = \exp\left(-\frac{1}{2\sigma^2}|x_i^p - c_a^p|^2\right). \tag{2}$$

assuming isotropic covariance matrices ($\Sigma_a^p = \sigma^2 I$), and, and, after normalization, $\sum_{a=1}^K m_{ai}^p = 1$. The temperature is defined to be the variance in equation 2, $T = \sigma^2$.

The algorithm proceeds in an iterative manner. The cluster center points are initiated at randomly chosen points in the dataset. They are recomputed after the membership has been updated (eq. 1), using

$$c_a^p = \frac{\sum_{i=1}^{N_p} m_{ai}^p x_i^p}{\sum_{i=1}^{N_p} m_{ai}^p}, \forall a \in \{1, \dots, K\}, \forall p \in \{1, \dots, P\}. \tag{3}$$

Then, the atlas points are formed by averaging the cluster center points, $z_a = 1/P \sum_{p=1}^P c_a^p$. The atlas points can be seen as a "common frame of reference", representing the mean shape of the point sets. An affine and non-rigid thin plate spline transform is now computed to warp the atlas points to the cluster center points and vice versa. The next iteration is started by lowering the temperature and transforming the atlas points to the subsequent datasets. In the latter step, information from other datasets is implicitly included in clustering a single dataset.

2.3 Statistics

DTI is sensitive to detecting pathology that manifests itself along white matter pathways. The challenge is to be precise in describing which local regions specifically characterize a disease. We assume for our analysis that the voxels through which a fiber tract passes are somehow connected. By jointly analyzing such samples, statistical power is to be gained compared to per-voxel comparison. What is more, a multi-modal analysis of the data may actually be necessary: simply averaging data over the connected voxels delivers a physiologically feasible mean only if there is a unimodal distribution.

We propose to use the clustering outcome of the registration as the statistical frame of reference; i.e. correspondence is implicitly defined via the atlas cluster points z_a. The membership function (equation 1) defines the extend to which

points x_i^p are assigned to cluster centers point c_a^p. It may be observed that in such a way fiber tract statistics are modeled in 3D since the data points emanate from several fibers that constitute a tract. To the best of our knowledge, previously proposed arc-length representations have not yet been generalized to higher dimensions. A cumulative distribution of FA-values per patient per cluster is build, in which the membership values act as relative weights. Consequently, distant voxels will get a low weight.

The clinical study described in this paper (see below) involves paired data of patients and controls. For each such pair, the distributions of FA data in the neighbourhood of a cluster center point is compared. In other words, a pair-wise comparison of FA-distributions is needed. As no prior information on the distribution is available, a non-parametric test is required. We use the Kolmogorov-Smirnov (KS)-test to do so in a two-tailed fashion: for each patient/control pair and for each cluster center point, it is determined if either the FA-distribution of the one is significantly above or below the other or if the difference is insignificant. The latter comparison is performed over all patient/control pairs from which the majority vote is retained. Effectively, this score boils down to either a higher or lower FA, ties are discarded.

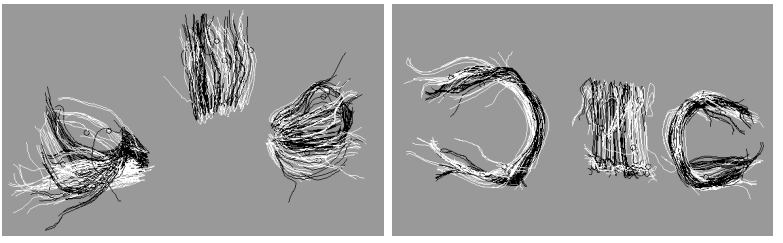
Subsequently, the consistency of the outcome along the tract is checked. Here, the underlying hypothesis is that if the voting score is similar over a larger number of nearby clusters, it is an indication of a significant difference between the groups. In order to check the 'consistency' two classes are asserted, corresponding to dominance of either the patient or control distributions. Class assignment is obtained by unsupervised classification by means of a k-nearest neighbour classifier. A leave-one(-cluster)-out cross validation scheme is involved to train and test the classifier, resulting in the relative classification error over all clusters. It is taken that a small classification error indicates that the dominating distribution can be predicted from neighbouring data. This is considered to reflect a systematic difference between the groups.

For comparison purposes a paired t-test is performed on averaged FA-values over the clusters, using the membership values as relative weights. Notice that such a comparison does not include any spatial connectivity.

To assess the validity of both approaches, statistics are computed on ten randomly permuted data sets (i.e. by randomly assigning the class labels 'patient' or 'control' to the data).

3 Results

Six infant survivors treated for medulloblastoma with intravenous methotrexate and cranial radiotherapy were included in the study as well as age-, education- and sex-matched healthy controls. The aim of this study was to confirm the hypothesis that the treatment affects a major white matter tract, the corpus callosum. The number of subjects emanated from a power analysis performed a priori, expecting a major effect. All subjects were scanned on a 3.0T Philips Intera (Best, The Netherlands) MRI-scanner. After data inspection, it was found



(a) sagittal (b) axial

Fig. 1. Transformed fibers of the corpus callosum (that could be reproducibly tracked in all subjects) after matching. Fibers of one patient and one control are shown in black and white.

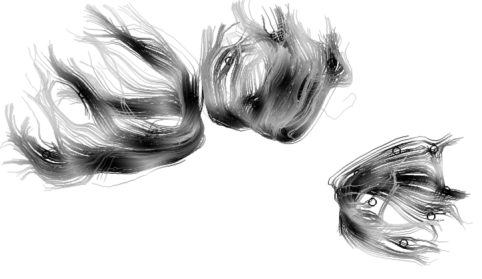


Fig. 2. The maximum membership per fiber-point $\max_i m_{ai}^p$ and cluster center points after one matching instance, in one dataset

that unfortunately one patient had incomplete data and had to be excluded (and simultaneously the matched control).

Fiber tracking was performed in the genu, center and splenium of the corpus callosum. Because of white matter degeneration, other corpus callosum parts could not be successfully tracked in all subjects. Fiber coordinates (in mm) were scaled by a factor of 0.01, such that increased numerical stability between translation and rotation parameters was obtained. The resulting tracts from the 10 datasets were jointly matched. The optimal cluster size is determined experimentally below (see figure 4). The temperature was step-wise lowered by a factor of 0.9, starting at $T_{init} = 10^{-2}$, until the final temperature of $T_{final} = 10^{-3}$ was reached. As an illustration, registered tracts of one randomly chosen patient/control pair are displayed in figure 1. To account for the inherent stochastic nature of the clustering, ten realizations of atlas points were combined.

Next, the membership m_{ai}^p was computed for all voxels included in the tracking. Here, $T = 5 \cdot 10^{-3}$ was chosen, yielding a uniform coverage of the clusters over the data. The result is depicted in figure 2. Subsequently, the cumulative FA-distributions were calculated using the membership function as weighting. In the Kolmogorov-Smirnov test to assess the difference per patient/control pair, the number degrees of freedom equaled the cluster size in voxels. The distributions in one cluster for all patients and corresponding KS-significances are given in figure 3.

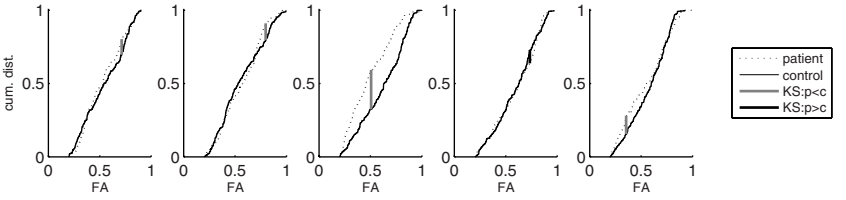


Fig. 3. Cumulative FA-distributions for five patient/control pairs in one arbitrarily chosen cluster, with Kolmogorov-Smirnov test results. Voting yields a lower FA for patients for this cluster.

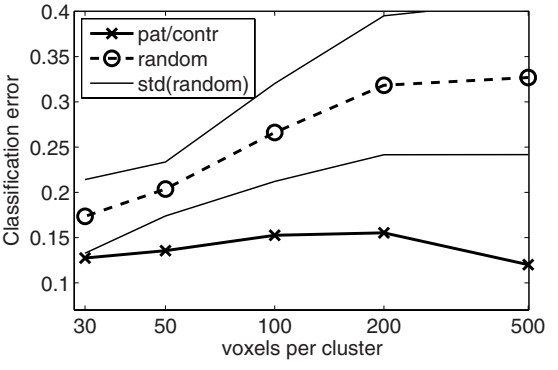
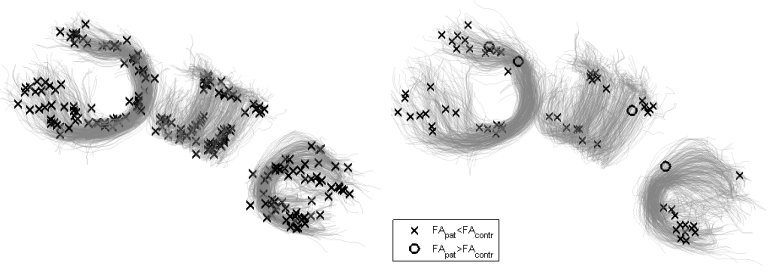


Fig. 4. Classification error as function of the cluster size for both patients/controls and randomly composed classes. The latter is a mean over ten different class compositions, for which the corresponding standard deviation is plotted in gray.

In a leave-one-out cross-validation, a $(3 \cdot 10)$-nearest neighbour classifier was trained on the voting outcome over all pairs of the 10 matching instances, and the classification error was determined. Registration and classification were executed ten times with a range of voxels per cluster. The results are shown in figure 4. The error for randomly composed classes is proportional to the cluster size. Due to the local continuity of the FA, neighbouring values can be predicted, regardless of the class composition. The difference with randomly assigned class labels was significant ($p<0.05$) for cluster sizes larger and 50 voxels [11]. The outcome of the voting and classification process for a cluster size of 200 voxels is displayed in figure 5. A globally decreased FA can be observed.

For comparison, a paired t-test per cluster center point was done (we did not include a correction for the number of comparisons at this stage). 22% of the total number of cluster center points yielded a significant difference ($p < 0.05$). These are annotated in figure 5. In comparison to the results of the 2-nn classifier, less consistent and more evenly distributed differing points can be observed, which we consider not trustworthy. Randomly permuting the class labels 500 times yielded that only in 4% of the permutations a significant difference was found



(a) 3-nn classifier (b) t-test

Fig. 5. Result of the voting and classification procedure, depicting correctly classified clusters with decreased (×) and increased FA (○) for patients. Also, the results of t-test statistics (without multiple comparison correction) are shown.

in 22% or more cluster center points. From this we conclude that the differences might be conceived as mathematically significant (since the percentage is lower than 5%), despite clinical reserves.

4 Discussion

We presented a novel approach to 3D, tract-based statistical analysis of FA data. Inter-subject correspondence was achieved by non-rigid registration based on a joint clustering and matching. The clustering delivered atlas points that served as a frame of reference for performing the analysis. A distribution of FA-values per cluster and per subject was calculated. Subsequently, the Kolmogorov-Smirnov test for inequality of distributions was used to pair-wise compare patients and controls. The latter comparison was performed over available patient/control pairs from which the majority vote was retained.

In a study into infant medulloblastoma survivors, corpus callosum tracts from patients and controls were pair-wise studied. A decreased FA in larger parts of the corpus callosum was observed. The decrease could be predicted by a 3-NN classifier with an error of 14% when 200 voxels per cluster were used, which indicated a significant difference. In comparison to t-test statistics per cluster, the proposed method was able to detect larger regions in which patients differed from controls.

Voxel-based analysis is well-known to be relatively sensitive to small misregistration. In contrast, the proposed method is merely affected by the misregistration component that is parallel to the tract. Due to the continuity of the FA, this effect could be limited. Actually, in figure 1, we illustrated a good overlap between tracts of an arbitrarily chosen patient/control pair.

Although manual tracking was performed in this study, we believe that this work can easily be adapted to (un)supervised full brain tractography clustering methods. Additionally, analyzing larger cohorts that are partitioned in several subgroups, could be facilitated by a multi-sample Kolmogorov-Smirnov test.

Acknowledgments

Matthan Caan is involved in the Virtual Laboratory for e-Science project, supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science, and the ICT innovation program of the Ministry of Economic Affairs. This study was financially supported by the Gratama Foundation and Research School Neurosciences Amsterdam, The Netherlands.

References

1. Basser, P., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Magn. Reson. B 111, 209–219 (1996)
2. Gerig, G., Gouttard, S., Corouge, I.: Analysis of brain white matter via fiber tract modeling. Engineering in Medicine and Biology Society 426, 4421–4424 (2004)
3. Sage, C., Peeters, R., Gorner, A., Robberecht, W., Sunaert, S.: Quantitative diffusion tensor imaging in amyotrophic lateral sclerosis. NeuroImage 34, 486–499 (2007)
4. Caan, M., Vermeer, K., van Vliet, L., Majoie, C., Peters, B., den Heeten, G., Vos, F.: Shaving diffusion tensor images in discriminant analysis: A study into schizophrenia. Medical Image Analysis 10, 841–849 (2006)
5. Corouge, I., Fletcher, P., Joshi, S., Gouttard, S., Gerig, G.: Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. Med. Im.Anal. 10, 786–798 (2006)
6. Maddah, M., Wells III, W., Warfield, S., Westin, C.F., Grimson, W.: Probabilistic clustering and quantitative analysis of white matter fiber tracts. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 372–383. Springer, Heidelberg (2007)
7. O'Donnell, L., Westin, C.F., Golby, A.: Tract-based morphometry. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 161–168. Springer, Heidelberg (2007)
8. Smith, S., Jenkinson, M., et al.: Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. NeuroImage 31, 1487–1505 (2006)
9. Jiang, H., van Zijl, P., et al.: DtiStudio: Resource program for diffusion tensor computation and fiber bundle tracking. Computer Methods and Programs in Biomedicine 81, 106–116 (2006)
10. Chui, H., Rangarajan, A., Zhang, J., Leonard, C.: Unsupervised learning of an atlas from unlabeled point-sets. IEEE Trans. on Pattern analysis and Machine Intelligence 26, 160–172 (2004)
11. Alippi, C., Braione, P.: Classification methods and inductive learning rules: What we learn from theory. IEEE Trans. Systems Man. Cybernatics 36, 649–655 (2006)